

**IN THE CLAIMS:**

**Kindly replace the claims with the following:**

1. (Currently amended) A method of calculating iteration values for free parameters

$\lambda_{\alpha}^{ortho(n)}$  of a maximum-entropy speech model MESM in a speech recognition system with the aid of the generalized iterative scaling training algorithm in accordance with the following formula:

$$\lambda_{\alpha}^{ortho(n+1)} = G(\lambda_{\alpha}^{ortho(n)}, m_{\alpha}^{ortho}, \dots)$$

where:

n : is an iteration parameter;

G : is a mathematical function;

$\alpha$  : is an attribute in the MESM; and

$m_{\alpha}^{ortho}$  : is a desired orthogonalized boundary value in the MESM for the attribute  $\alpha$ ,

characterized in that the desired orthogonalized boundary value  $m_{\alpha}^{ortho}$  is calculated by

linearly combining the desired boundary value  $m_{\alpha}$  with desired boundary values  $m_{\beta}$  of attributes  $\beta$  that have a larger range than the attribute  $\alpha$ .

2. (Currently amended) A method as claimed in claim 1, characterized in that the

calculation of the desired orthogonalized boundary value  $m_{\alpha}^{ortho}$  for the attribute  $\alpha = [[\beta_0]]$  comprises the following steps:

- a) Selecting all the attributes  $[[\beta_i]]$   $\beta_i$  with  $i=1\dots g$  in the speech model that have a larger range RW than the attribute  $\alpha = [[\beta_0]]$   $\beta_0$  and include the latter;
- b) Calculating desired boundary values  $m_{\beta_i}$  for the attributes  $[[\beta_i]]$   $\beta_i$  with  $i=0\dots g$ ;
- c) Sorting the attributes  $[[\beta_i]]$   $\beta_i$  with  $i=0\dots g$  according to their RW;
- d) Selecting one of the attributes  $[[\beta_i]]$   $\beta_i$  having the largest RW;

- e) Checking whether there are other attributes  $[[\beta k]] \underline{\beta}_k$  which include the attribute  $[[\beta i]] \underline{\beta}_i$  and have a larger RW than the selected attribute  $[[\beta i]] \underline{\beta}_i$ ;
- f1) If so, defining a parameter X as a linear combination of the orthogonalized boundary values  $m_{\beta k}^{ortho}$  calculated in step g) during the last run of the steps e) to g) for all the attributes  $[[\beta k]] \underline{\beta}_k$  that have a larger range and are determined in the most recently run step e);
- f2) If not, defining the parameters X to  $X = 0$ ;
- g) Calculating the desired orthogonalized boundary value  $m_{\beta i}^{ortho}$  for the attribute  $[[\beta i]] \underline{\beta}_i$  by arithmetically combining the desired boundary value  $[[m\beta i]] \underline{m\beta}_i$  with a parameter X; and
- h) Repeating the steps e) to g) for the attribute  $[[\beta i]] \underline{\beta}_i - 1$  whose RW is smaller than or equal to the RW of the attribute  $\beta i$  until the desired orthogonalized boundary value  $m_{\beta 0}^{ortho} = m_{\alpha}^{ortho}$  with  $i=0$  has been calculated in step g).

3. (Original) A method as claimed in claim 2, characterized in that the calculation of the parameter X in step f1) is made according to the following formula:

$$X = \sum_k m_{\beta k}^{ortho}$$

4. (Original) A method as claimed in claim 3, characterized in that the calculation of the desired orthogonalized boundary value  $m_{\beta i}^{ortho}$  is made in step g) according to the following formula:

$$m_{\beta i}^{ortho} = m_{\beta i} - X$$

5. (Currently amended) A method as claimed in claim 2, characterized in that the calculation of the desired boundary values  $m_{\beta i}$  for the attributes  $[[\beta i]] \underline{\beta}_i$  with  $i = 0, \dots, g$  is

made in step b) by respectively calculating the frequency  $N(\beta_i)$ , with which the attribute  $[[\beta_i]] \underline{\beta}_i$  occurs in a training corpus and by subsequently smoothing the calculated frequency value  $N([[ \beta_i]] \underline{\beta}_i)$ .

6. (Currently amended) A method as claimed in claim 5, characterized in that the calculation of the frequency  $N(\beta_i)$  is made by applying a binary attribute function  $[[f\beta_i]] f\beta_i$  to the training corpus where  $[[f\beta_i]] f\beta_i$  is defined as:

$$f_{\beta_i}(h, w) f_{\beta_i}(h, w) = \begin{cases} 1 & \text{if } \beta_i \text{ fits in the word sequence } (h, w) \\ 0 & \text{otherwise} \end{cases}$$

and where  $f_{\beta_i}(h, w)$  indicates whether the attribute  $[[\beta_i]] \underline{\beta}_i$  correctly describes a pattern predefined by the word sequence  $(h, w)$ .

7. (Original) A method as claimed in claim 1, characterized in that the mathematical function  $G$  has as a further variable the magnitude of a convergence step  $t_{\alpha}^{ortho}$  with:

$$t_{\alpha}^{ortho} = 1/M^{ortho}$$

where

$M^{ortho}$ : represents for binary functions  $f_{\alpha}^{ortho}$  the maximum number of functions which yield the value 1 for the same argument  $(h, w)$ .

8. (Original) A method as claimed in claim 7, characterized in that the attribute function  $f_{\alpha}^{ortho}$  is calculated by linearly combining an attribute function  $f_{\alpha}$  with orthogonalized attribute functions  $f_{\beta}^{ortho}$  is calculated from attributes  $\beta$  that have a larger range than the attribute  $\alpha$ .

9. (Currently amended) A method as claimed in claim 8, characterized in that the calculation of the orthogonalized attribute function  $f_{\alpha}^{ortho}$  for the attribute  $\alpha = [\beta_0]$  comprises the following steps:
- a) Selecting all the attributes  $[\beta_i]$   $\beta_i$  with  $i=1 \dots g$  in the speech model that have a larger range RW than the attribute  $\alpha = [\beta_0]$   $\beta_0$  and include the latter;
  - b) Calculating boundary values  $[f\beta_i]$   $f\beta_i$  for the attributes  $[\beta_i]$   $\beta_i$  with  $i=0 \dots g$ ;
  - c) Sorting the attributes  $[\beta_i]$   $\beta_i$  with  $i=0 \dots g$  according to their RW;
  - d) Selecting one of the attributes  $\beta_i$  having the largest RW;
  - e) Checking whether there are other attributes  $\beta_k$  which include the attribute  $[\beta_i]$   $\beta_i$  and have a larger RW than the selected attribute  $[\beta_i]$   $\beta_i$ ;
  - f1) If so, defining a function F as a linear combination of the orthogonalized attribute function  $f_{\beta_k}^{ortho}$  calculated in step g) during the last run of the steps e) to g) for all the attributes  $[\beta_k]$   $\beta_k$  that have a larger range determined in the most recently run step e);
  - f2) If not, defining the function F to  $F = 0$ ;
  - g) Calculating the orthogonalized attribute function  $f_{\beta_i}^{ortho}$  for the attribute  $\beta_i$  by arithmetically combining the attribute function  $[f\beta_i]$   $f\beta_i$  with the function F; and
  - h) Repeating the steps e) to g) for the attribute  $\beta_{i-1}$  whose range is smaller than or equal to the range of the attribute  $[\beta_i]$   $\beta_i$  until the orthogonalized attribute function  $f_{\beta_0}^{ortho} = f_{\alpha}^{ortho}$  with  $i=0$  has been calculated in step g).

10. (Original) A method as claimed in claim 9, characterized in that the calculation of the function F in step f1) is made according to the following formula:

$$F = \sum_k f_{\beta_k}^{ortho}$$

11. (Original) A method as claimed in claim 9, characterized in that the calculation of the orthogonalized attribute function  $f_{\beta i}^{ortho}$  in step g) is made according to the following formula:

$$f_{\beta i}^{ortho} = f_{\beta i} - F$$

12. (Original) A method as claimed in claim 1, characterized in that the mathematical function G has the following form:

$$\begin{aligned} \lambda_{\alpha}^{ortho(n+1)} &= G(\lambda_{\alpha}^{ortho(n)}, m_{\alpha}^{ortho}, \dots) \\ &= \lambda_{\alpha}^{ortho(n)} + t_{\alpha}^{ortho} \cdot \log \left( \frac{[t_{\alpha}^{ortho} \cdot m_{\alpha}^{ortho} + b_{\alpha}]}{[t_{\alpha}^{ortho} \cdot m_{\alpha}^{ortho(n)} + b_{\alpha}]} \cdot \frac{1 - \sum_{\gamma} [t_{\gamma}^{ortho} \cdot m_{\gamma}^{ortho(n)} + b_{\gamma}]}{1 - \sum_{\gamma} [t_{\gamma}^{ortho} \cdot m_{\gamma}^{ortho} + b_{\gamma}]} \right) \end{aligned}$$

where:

$\alpha$  : refers to a just considered attribute;

$\gamma$  : refers to all the attributes in the speech model;

$t_{\alpha}^{ortho}, t_{\gamma}^{ortho}$  : refer to the size of the convergence step with  $t_{\alpha}^{ortho} = t_{\gamma}^{ortho} = 1/M^{ortho}$  with

$$M^{ortho} = \max_{(h,w)} \left( \sum_{\beta} f_{\beta}^{ortho}(h, w) \right);$$

where  $M^{ortho}$  for binary functions  $f_{\beta}^{ortho}$  represents the maximum number of functions which yield the value 1 for the same

argument

(h,w);

$m_{\alpha}^{ortho}, m_{\gamma}^{ortho}$  : refers to desired orthogonalized boundary values in the MESM for the attributes  $\alpha$  and  $\gamma$ ;

$m_{\alpha}^{ortho(n)}, m_{\gamma}^{ortho(n)}$  : refers to iterative approximate values for the desired boundary values  $m_{\alpha}^{ortho}, m_{\gamma}^{ortho(n)}$ ; and  
 $b_{\alpha}$  and  $b_{\gamma}$  : refer to constants.

13. (Currently amended) A method as claimed in claim 1, characterized in that the mathematical function has the following form:

$$\lambda_{\alpha}^{ortho(n+1)} = G(\lambda_{\alpha}^{ortho(n)}, m_{\alpha}^{ortho}, \dots)$$

$$= \lambda_{\alpha}^{ortho(n)} + t_{\alpha}^{ortho} \cdot \log \left( \frac{m_{\alpha}^{ortho}}{m_{\alpha}^{ortho(n)}} \cdot \frac{1 - \sum_{\beta \in Ai(n)} (t_{\beta} \cdot m_{\beta}^{ortho(n)})}{1 - \sum_{\beta \in Ai(n)} (t_{\beta} \cdot m_{\beta}^{ortho})} \right)$$

where:

$n$  : represents the iteration parameter;

$[[Ai(n)]] \underline{A}_{i(n)}$  : represents an attribute group  $[[Ai(n)]] \underline{A}_{i(n)}$  with  $1 \leq i \leq m$  selected in the  $n^{th}$  iteration step;

$\alpha$  : represents a just considered attribute from the just selected attribute group  $[[Ai(n)]] \underline{A}_{i(n)}$ ;

$\beta$  : represents all the attributes of the attribute group  $Ai(n)$ ;

$t_{\alpha}^{ortho}, t_{\beta}^{ortho}$  : represents the size of a convergence step with  $t_{\alpha}^{ortho} = t_{\beta}^{ortho} = 1/M_{i(n)}^{ortho}$

with

$$M_{i(n)}^{ortho} = \max_{(h,w)} \left( \sum_{\beta \in Ai(n)} f_{\beta}^{ortho}(h, w) \right)$$

where  $M_{i(n)}^{ortho}$  represents for binary functions  $f_{\beta}^{ortho}$  the maximum number of functions from the attribute group  $[[Ai(n)]] \underline{A}_{i(n)}$  which yield the value 1 for the same argument  $(h, w)$ ;

$m_{\alpha}^{ortho}$ ,  $m_{\beta}^{ortho}$  : represent desired orthogonalized boundary values in the MESM for the attributes  $\alpha$  and  $\beta$  respectively;

$m_{\alpha}^{ortho(n)}$ ,  $m_{\beta}^{ortho(n)}$  : represent iterative approximate values for the desired boundary values  
 $m_{\alpha}^{ortho}$ ,  $m_{\beta}^{ortho}$  ;

where the selection of the group  $[[A_i(n)]]$   $A_{i(n)}$  of attributes  $\alpha$ , whose associated parameters  $\lambda_{\alpha}^{ortho}$  are adapted to a current iteration step is made either cyclically or according to a predefined criterion.

14. (Original) A speech recognition system (10) comprising: a recognition device (12) for recognizing the semantic content of an acoustic signal captured and rendered available by a microphone (20), more particularly a speech signal, by mapping parts of this signal onto predefined recognition symbols as they are offered by the implemented maximum-entropy speech model MESM, and for generating output signals which represent the recognized semantic content; and a training system (14) for adapting the MESM to recurrent statistical patterns in the speech of a certain user of the speech recognition system (10); characterized in that the training system (14) calculates free parameters  $\lambda$  in the MESM in accordance with the method as claimed in claim 1.

15. (Original) A training system (14) for adapting the maximum-entropy speech model MESM in a speech recognition system (10) to recurrent statistical patterns in the speech of a certain user of this speech recognition system (10), characterized in that the training system (14) calculates free parameters  $\lambda$  in the MESM in accordance with the method as claimed in claim 1.